DOCUMENT RESUME

ED 307 337 TM 013 521

AUTHOR Rosa, Carlos M.; And Others
TITLE A Comparison of Two Standardized Reading and Mathematics Achievement Tests in the Native Language for Hispanic Limited-English-Proficient Students.
PUB DATE Mar 89
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Bilingual Education; Comparative Analysis; Cultural Background; Difficulty Level; Elementary Education; Elementary School Students; Hispanic Americans; *Limited English Speaking; *Mathematics Tests; Psychometrics; *Reading Tests; Spanish Speaking; *Standardized Tests; Teacher Attitudes; Test Format; Test Items; Test Reliability; Test Validity
IDENTIFIERS *Prueba Riverside de Realizacion en Espanol

ABSTRACT

A study was undertaken to examine psychometric properties of "La Prueba Riverside de Realizacion en Espanol" (PRRE) and the "Spanish Assessment of Basic Education" (SABE) when administered to a sample of limited-English-proficient students, grades 1 through 8. Spanish-language versions of both tests were used for the study. Subjects included a total of 2,634 students who were enrolled in 19 elementary schools in Chicago and who participated in a Spanish bilingual education program. The instruments, both of which include reading and mathematics subtests, were administered by bilingual teachers who received inservice training to prepare them for their part in the study. In addition to analyses of test results, questionnaire data from participating teachers were assessed. The teacher questionnaire covered seven areas--item difficulty, test instructions, test fit with the curriculum, language correctness and appropriateness, racial or ethnic test bias, print size and illustrations, and cultural relevance of items. Results indicate that the PRRE and SABE are acceptable for the population assessed. The instruments are essentially similar in terms of psychometric properties and in teachers' ratings of test characteristics. Nine data tables are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED307337

A COMPARISON OF TWO STANDARDIZED READING AND MATHEMATICS ACHIEVEMENT TESTS IN THE NATIVE LANGUAGE FOR HISPANIC LIMITED-ENGLISH-PROFICIENT STUDENTS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CARLOS M. ROSA

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Carlos M. Rosa
Carole L. Perlman
Arie van der Ploeg

Chicago Public Schools
Department of Research and Evaluation

Paper prepared for presentation at the annual meeting of the
American Educational Research Association
San Francisco, California, March 1989

TM013521

BEST COPY AVAILABLE

TABLE OF CONTENTS

	Page
INTRODUCTION	1
METHODOLOGY	2
STUDENT CHARACTERISTICS	3
DESCRIPTION OF INSTRUMENTS	6
TEST ADMINISTRATOR QUESTIONNAIRE RESULTS	9
LA PRUEBA AND SABE TEST RESULTS	12
CONCLUSIONS	16
REFERENCES	17

List of Tables

Table		Page
1	Students' Ethnic Background.....	3
2	Number and Percent of Students by Instructional Category and Grade Level.....	5
3	La Prueba and SABE Levels by Grade Ranges.....	7
4	SABE-Approximate Testing Time in Minutes.....	8
5	La Prueba-Approximate Testing Time in Minutes.	8
6	Mean Teacher Ratings by Test Subject Area.....	11
7	Comparison of SABE and La Prueba Reading and Math Results by Grade Level.....	12
8	Comparison of SABE and La Prueba Reading and Math Results by Student Bilingual Instructional Category.....	13
9	La Prueba and SABE Reliability and SEM for Grades 3,6 and 8.....	15

A Comparison of Two Standardized Reading and Mathematics Achievement Tests in the Native Language for Hispanic Limited-English-Proficient Students

Introduction

The use of standardized, commercially developed, English-language achievement tests for limited-English-proficient (LEP) students has serious drawbacks. The students' lack of facility with English impedes their performance, making it difficult to obtain an accurate assessment of their skills. For students who receive the majority of their instruction in their native language, it has been suggested that, whenever such tests exist in the primary language, they can and should be used (Piper, 1987). Piper, Doherty, and Russo (1982) have documented the degree to which Spanish-dominant LEP students perform better on the Spanish reading and language subtests of the CTBS Español than they do on the English reading and language subtests of the CTBS, Form S. The relationship between native language instruction and native language test results is well known, yet some educators continue to use English language reading and language subtests of norm-referenced batteries with LEP students for whom such a test will yield results of questionable validity. One alternative is to administer achievement tests to LEP bilingual program participants in their native language. The purpose of this paper is to describe a study of two such achievement batteries.

During the months of May and June 1988 the Chicago Public Schools (Department of Research and Evaluation) conducted a field-test of the two well known and most widely used Spanish language achievement tests. The two instruments selected for

this study were: La Prueba Riverside de Realización en Español (La Prueba), published by Riverside Publishing Company, and the Spanish Assessment of Basic Education (SABE), published by CTB/McGraw Hill.

Methodology

The purpose of this field test was to examine the psychometric properties of these two tests when administered to a sample of limited-English-proficient (LEP) students participating in the Spanish bilingual education program in Chicago public elementary schools. Test reliabilities and standard errors of measurement were calculated. Content validity was evaluated based on a survey completed by teachers who administered both tests. The results of the study were intended to help determine which test to recommend for use with Spanish bilingual program students on a citywide basis.

Nineteen elementary schools participated in this field-testing. The 19 schools were chosen on the basis of their geographical location within the city and their student characteristics. A total of 2,634 limited-English-proficient students in grades 1 through 8 were administered both La Prueba and the SABE. The sample included Hispanic students from Cuba, Mexico, Puerto Rico and Central America (Table 1).

An inservice training session was conducted for the teachers coordinating the field testing at each school. Topics included test administration procedures for La Prueba and the SABE tests and the necessity to maintain uniformity in the test administration procedures. A manual outlining the testing

procedures was provided to all the teachers involved in the testing project.

Table 1

Students' Ethnic Background

Ethnic Group	Number	Percent
Mexican	1,897	72
Puerto Rican	478	18
Cuban, Central American and Other	259	10
Total	2,634	100

Student Characteristics

The Chicago public school system classifies limited-English-proficient students into four instructional categories for bilingual education program placement. The instructional categories are determined through evaluation of the student's English language proficiency using a locally developed standardized English fluency test and teacher's evaluation of the students' English language fluency. Students placed into Category A have very little or no understanding of English and receive almost all their instruction in the native language. Students in Category B speak and understand some English and receive half of their instruction in the native language. Students in Category C speak and understand English well enough to participate in a classroom in which English is

used most of the time and receive almost all their instruction in English. English language proficiency in category GP (General Program) is at a level needed to perform adequately in an all-English classroom; these students receive all their instruction in English.

Most of the LEP students, 47 percent in the field-test sample, fell in instructional Category A (Table 2). Most of the students in Category A were recent arrivals to the United States. These were generally in their first year in the bilingual education program. They were concentrated in the lowest grades. Thirty-three percent of the sample comprised Category B students. The majority of category B students had been enrolled in the bilingual education program for two years. Category C students generally receive bilingual services for about three years and most of them were concentrated in grades 4-6. Category GP indicates a student who is in the general program of instruction (as opposed to bilingual education); there were no Category GP students included in the study.

Most of the students in the field test sample were dominant in the Spanish language, received most of their instruction in Spanish, had received bilingual services from 1 to 3 years, and were enrolled in grades 1-6.

Table 2

Number and Percent of Students
by Instructional Category and Grade Level

Instructional Category*	Grade Level 1	2	3	4	5	6	7	8	Total
A	494 18.8	321 12.2	125 4.8	7 2.9	90 3.4	45 1.7	53 2.0	34 1.3	1239 47.0
B	51 1.9	173 6.6	176 6.7	159 6.0	102 3.9	86 3.3	70 2.7	60 2.3	877 33.3
C	3 0.1	22 0.8	64 2.4	117 4.4	84 3.2	87 3.3	81 3.1	60 2.3	518 19.7

* Instructional category A students have no or little understanding of English, category B understand some English, and category C are proficient in English but not well enough to participate in the general program of instruction.

Description of Instruments

La Prueba consists of separate levels of tests for grades K through 8, with level 6 corresponding to Kindergarten and level 14 corresponding to grade 8. It includes reading, mathematics, science and social studies subtests. Reading and mathematics are tested at all levels. Language is measured at levels 9-14 (grades 3-8) and science and social studies are measured at levels 8-14 (grades 2-8). La Prueba was normed on a sample of Spanish-speaking students in Texas; the publisher refers to these students as the Texas Reference Groups.

The SABE is a Spanish-language reading and math achievement test designed for grades 1 through 8. The reading subtest measures three areas: word attack, reading vocabulary, and reading comprehension. The mathematics subtest provides 2 measures, one of mathematics computation, and one of mathematics concepts and applications. SABE has six levels that overlap grades 1 through 8. The norms for the test were developed from tryout data collected on bilingual program students throughout the United States.

The levels and grade ranges for La Prueba and SABE tests are given in Table 3.

Table 3

La Prueba and SABE Levels and Grade Ranges

Grade	La Prueba Level	SABE Level
1	7	1
2	8	2
3	9	3
4	10	4
5	11	5
6	12	5
7	13	6
8	14	6

Tables 4 & 5 indicate the approximate working times per grade for each of the instruments.

Table 4

SABE - Approximate Testing Time in Minutes

Grade	Word Attack	Vocabulary	Reading Compre-hension	Computation	Concepts and Applications
1	27	37	35	20	24
2	35	19	28	18	34
3	22	30	36	34	33
4	--	29	45	33	37
5	--	29	45	33	37
6	--	29	45	33	37
7	--	29	45	33	37
8	--	29	45	33	37

Table 5

LA PRUEBA
Approximate Testing Time in Minutes

Grade	Reading	Language Arts	Math
1	30	--	30
2	30	--	30
3	30	--	30
4	35	25	30
5	45	25	30
6	45	25	30
7	45	25	30
8	45	25	30

Test Administrator Questionnaire Results

At the conclusion of the field testing, the bilingual teachers who administered the instruments completed a questionnaire about the administration of the tests as well as their opinions of the tests' content validity. Sixty-eight percent of the 124 teachers responded to the questionnaire.

Teachers were asked to rate the SABE and La Prueba. Seven test characteristics were offered:

- The level of item difficulty seems appropriate for the grade.
- The instructions are clear and appropriate for the grade.
- The test correlates to the curriculum being used in your school.
- The Spanish used is appropriate.
- The test contains no racial/ethnic biases.
- The test has adequate size print and illustrations.
- The items are culturally relevant.

Teachers were asked to specify their agreement with each characteristics, for each subtest of each test, using the following forced-choice scale: 1=strongly disagree, 2=disagree, 3=agree, and 4=strongly agree.

Since the tests have different subtest structures, teachers' responses for the SABE's Fonética and Vocabulario subtests were averaged to produce a language mean to compare to La Prueba's Language subtest. The same was done for the SABE's two mathematics subtests. Three sets of comparisons are therefore possible: in reading comprehension, language skills, and mathematics. Table 6 presents the mean ratings, the number of

respondents, the correlation between the SABE and La Prueba ratings, and the (paired) t-statistics.

The teachers generally agreed that the language and mathematics subtests for both instruments were reasonable: almost all mean ratings exceeded the 3.0 (agree) level. Agreement on the reading subtests was somewhat lower although no mean rating fell below 2.5, the mid-point on these forced four-choice ratings, confirming that more selected the agree categories than the disagree. Item difficulty and curriculum match were viewed by these teachers as providing the most cause for concern. In general, it would appear that the teachers found both tests acceptable. Their mean ratings did not vary significantly between the two tests. On the other hand, the correlations between the teachers' ratings of La Prueba and SABE subtests generally ranged between about 0.5 and 0.7, suggesting that about one-half to three-quarters of the variance in the ratings was unique to each test.

Mean Teacher Ratings by Test Subject Area

	Reading					Language					Mathematics				
Criterion:	SABE	Prueba	N	r	t*	SABE	Prueba	N	r	t*	SABE	Prueba	N	r	t*
Item difficulty Appropriate	2.75	2.93	56	0.51	-1.35	3.10	2.98	56	0.52	1.01	3.21	3.02	56	0.34	1.47
Quality of Instruction	3.15	3.05	57	0.61	1.19	3.22	3.04	57	0.62	1.74	3.27	3.05	57	0.43	1.78
Match with local curriculum	2.87	2.87	54	0.48	0.00	2.98	2.88	57	0.57	0.98	3.04	2.98	55	0.43	1.24
Spanish appropriate	2.98	3.02	56	0.67	-0.39	3.16	3.12	57	0.71	0.43	3.16	3.07	56	0.66	0.99
Racial/ethnic bias	3.24	3.15	53	0.71	1.04	3.27	3.15	54	0.57	1.48	3.28	3.18	54	0.75	1.14
Print & illustrations	3.11	3.07	55	0.66	0.38	3.23	3.11	56	0.63	1.34	3.29	3.12	56	0.52	1.60
Cultural relevance	2.96	2.98	45	0.74	-0.26	3.04	3.04	45	0.84	0.00	3.15	3.04	44	0.69	0.88

*The critical region, alpha = .05, two tailed, falls beyond 2.00.

La Prueba and SABE Test Results

Table 7 presents the basic reading and mathematics results, using percent correct scoring by grade for both La Prueba and SABE tests. The La Prueba reading results are slightly higher than the SABE scores, while the mathematics results are essentially equal.

Table 7

Comparison of SABE and La Prueba Reading and Math Results by Grade Level (using percentage correct scoring)

	Reading			Mathematics		
Grade	N	SABE	Prueba	N	SABE	Prueba
1	524	68.2	75.4	523	72.7	73.4
2	510	61.2	66.1	508	72.3	65.5
3	371	61.1	60.3	366	63.2	59.5
4	344	56.9	62.4	345	65.2	65.3
5	276	51.8	62.0	276	49.3	56.6
6	247	55.9	57.6	244	57.3	56.2
7	206	53.2	50.0	199	47.9	55.0
8	157	60.0	57.7	156	57.3	57.7

Table 8

Comparison of SABE and La Prueba Reading and Math Results by Student Bilingual Instructional Category

Bilingual Instructional Category	Reading			Mathematics		
	N	SABE	Prueba	N	SABE	Prueba
A	1,204	62.1	66.2	1,199	65.3	65.0
B	870	58.7	62.4	863	62.4	61.1
C	514	56.7	60.1	508	61.6	62.5

Notes: The SABE Reading Composite is compared to the La Prueba Reading subtest; the SABE Mathematics Composite is compared to the La Prueba Mathematics subtest. These are the best matches in terms of test domain.

Both tests are somewhat too easy for this population at the first and possibly the second grade levels. At the other grades, however, there was no evidence of either ceiling or floor effects. The results were also analyzed by English-language ability of the population (Table 8). As would be expected of well-made tests, no major variations were found.

To examine the internal consistency of both tests, the Kuder-Richardson Formula 20 reliabilities (KR-20) and the standard errors of measurement (SEM) were calculated. These indices are displayed in Table 9. The KR-20 and SEM results shown are for students in grades 3,6 and 8. The data reveal that both tests have a high internal consistency.

The SABE reliabilities are slightly higher at all grade levels for reading and mathematics. This is probably a function of the SABE test having more items per test level than La Prueba test. The standard error of measurement is slightly smaller for La Prueba than the SABE, a fact again explained by the larger number of items of the SABE test.

Table 9

La Prueba and SABE
Reliability and SEM for Grades 3,6 and 8.

Reading

Grade 3

	La Prueba	SABE
Reliability (KR-20)	.813	.830
SEM	2.514	2.548

Grade 6

	La Prueba	SABE
Reliability (KR-20)	.779	.901
SEM	2.541	3.102

Grade 8

	La Prueba	SABE
Reliability (KR-20)	.818	.861
SEM	2.569	3.144

Mathematics

Grade 3

	La Prueba	SABE
Reliability (KR-20)	.796	.837
SEM	2.550	3.001

Grade 6

	La Prueba	SABE
Reliability (KR-20)	.767	.897
SEM	2.596	3.092

Grade 8

	La Prueba	SABE
Reliability (KR-20)	.719	.864
SEM	2.609	3.176

Conclusions

The field-testing program for La Prueba and SABE Tests has demonstrated that both instruments are acceptable for the population in question. The SABE and La Prueba are essentially similar in terms of psychometric properties and in teachers' ratings of test characteristics. This study explicitly avoided the test publishers' normative scores. In one case the normative sample is unacceptably small and insular. In the other case norming and equating procedures seem unnecessarily complex. Regardless of which test is chosen, local school district norms have to be developed.

The field test was to determine if either test was unacceptable and, some administrators hoped, to provide a rationale for the final decision to purchase. This did not occur. Because both tests were psychometrically sound, the final recommendation was based on issues not related to psychometric properties and content validity. Other statistical techniques such as Rasch analysis are currently in process and may provide more clear distinctions between the two tests' psychometric properties.

Any school system choosing a Spanish-language assessment battery should pay close attention to the match with local school curriculum, applicability of norms, and the availability of desired subtests. If instruments are similar in content and students test similarly, other factors should be considered in the selection such as: test format, grade level or functional level testing, time of administration, scores provided, and the cost of the test.

References

Cardenas, J.A. (1986). The Role of Native-Language Instruction in Bilingual Education. Phi Delta Kappan, 359-363.

CTB/McGraw-Hill. 1987. Spanish Assessment of Basic Education. Monterey, CA: CTB/McGraw-Hill.

Illinois Resource Center. 1985. Assessment of Language Minority Students, A Handbook for Educators. Arlington Heights, Ill: IRC.

The Riverside Publishing Company. 1984. La Prueba Riverside de Realización en Español. Chicago, Ill: The Riverside Publishing Company.

Piper, R. (1987). Effective Bilingual Education Evaluation: Is it Possible? Issues of Language Assessment, Vol. 3: Language Assessment and Public Policy, 79-91.

Wargo, M.J. & Green, D.R. (1977). Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation. Monterey, CA: CTB/McGraw-Hill.